GEORGIA M. PESTANA
Corporation Counsel

THE CITY OF NEW YORK
LAW DEPARTMENT
100 CHURCH STREET
NEW YORK, NEW YORK 10007

ANTHONY DISENSO
Senior Counsel
adisenso@law.nyc.gov
(212) 356-2640

September 17, 2021

By ECF
Honorable Gabriel W. Gorenstein
United States Magistrate Judge
Southern District of New York
500 Pearl Street
New York, New York 10007

      In Re: *New York City Policing During Summer 2020 Demonstrations*,
      No. 20 Civ. 8924 (CM) (GWG)
      This filing is related to all cases

Your Honor:

      I am an E-Discovery Counsel in the office of Georgia M. Pestana, Corporation Counsel of the City of New York, and one of the attorneys representing the defendants in the above-referenced matter for matters related to electronically stored information. I write in response to the Court's Order dated September 14, 2021 to provide additional information on the City's email collection and review process for Office of the Mayor and NYPD email. (*See*, Dkt. No. 263).

Email Collection Effort

      On August 6th, the parties largely reached agreement on the collection criteria for defendants' 49 agreed-upon custodians.[1] Throughout the month of August, the Law Department worked with technical staff at the Office of the Mayor and the NYPD to collect the email in an efficient and defensible manner. Despite the large volume of custodians and resulting data size, the Office of the Mayor's collection was completed on August 16th and the NYPD's was completed on September 1st. Combined, the collections totaled approximately 1.2 million documents and almost a terabyte of data.

Email Processing and Data Assessment

      Once the Law Department received the collections, they were run through quality control to ensure that the collection criteria was properly applied. The data was then sent to our vendor, Epiq to be processed and uploaded into a review platform. Processing included de-duplicating both sets of documents, which reduced the population from 1.2 million documents to

---

[1] By email dated August 6th, defendants accepted plaintiffs additional proposed custodians and search terms with two limited exceptions: (i) one term contained unsearchable characters, and (ii) a term related to the January 6th Insurrection would not have yielded responsive material as the time period for collection ended prior to the election.

approximately 460,000 documents. Defendants then analyzed a metadata export of the collections to identify documents that did not need full review and were able to reduce the review population by approximately 82,000 documents.[2] The email was then threaded to identify the most-inclusive email in the review set. This reduced the review population to approximately 286,000 documents.

Defendants then reviewed statistically valid samples of 400 documents from each of the email sets to determine the most efficient review approach. The sample review also allowed defendants to provide a rough estimate of the percentage of documents in each population that were likely to be responsive and identify technology assisted review (TAR) workflows that could potentially reduce the review burden to under 100,000 documents.

The time spent over the last two weeks following this iterative process has allowed defendants to potentially reduce the review burden by as much as 90% without significant cost to the City or added time to the email review.

Timeframe for Email Review Completion

At this point, defendants estimate that review and production would take approximately 11 weeks to complete working with a managed review team of 20 attorneys.[3] This estimate includes the time necessary to hire and train a review team. Many factors could impact this estimate including required additional quality control, review of additional documents identified by the TAR process, and the production schedule.[4] For example, if plaintiffs insist on frequent rolling productions, resources must be diverted from review to finalize each production set, extending the total time needed to complete the review. Additionally, we continue to work with our clients to identify an alternative workflow to deal with information that is particularly sensitive from a law enforcement perspective. This too may impact the time needed to complete the review.

---

[2] Defendants took this step after discovering many documents not needing review in an initial sample.

[3] The Law Department anticipates utilizing a continuous active learning (CAL) workflow here. CAL continuously reorders documents in a review queue to prioritize likely responsive documents as the software refines its understanding of responsiveness. Review continues until several review rounds include few, if any, marginally responsive documents.   At that point, the validation process starts by taking a statistically valid sample of the remaining unreviewed documents. Depending on the results, the review is either validated or additional review is done.

[4] Defendants note that this estimate relates specifically to agency email review. Plaintiffs have demanded that defendants collect and produce from several different data sources for each custodian. We are in the process of conducting custodial interviews to identify any additional sources that should be collected and added to the review.

Defendants recognize that this timeline presents challenges given the current discovery schedule but note that it is extremely aggressive given the extraordinary data collected from the 49 custodians demanded by plaintiffs and needing review. Defendants are currently in the process of hiring managed review team and anticipate beginning review the week of September 27th.[5]

Respectfully submitted,

Anthony DiSenso s/

Anthony DiSenso
Senior Counsel
E-Discovery Division

cc: ALL COUNSEL (via ECF only)

---

[5] Once email review and production is completed, defendants will immediately begin work on a privilege log to the extent necessary.